Linear Regression
Model: $f : \mathbf{x} \rightarrow y, \;\; f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$

Training data: $\{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_N}, y_N)\}$, or $\mathbf{X} = \begin{bmatrix} -\ \mathbf{x_1}^T\ - \\ \vdots \\ -\ \mathbf{x_N}^T\ - \end{bmatrix}$ and $\mathbf{y} = [y_1, \ldots, y_N]^T$

Evaluation through residual sum of squares (no regularization):
$J(\mathbf{w}) = RSS(\mathbf{w}) = \sum_{n=1}^{N} (y_n - \mathbf{w}^T\mathbf{x_n})^2 = (\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w})$

Evaluation through residual sum of squares ($l2$-norm regularization):

$$\begin{aligned} J(\mathbf{w}) &= RSS(\mathbf{w}) + \lambda \sum_{d=1}^{D} w_d^2 \\ &= RSS(\mathbf{w}) + \lambda \|\mathbf{w}\|_2^2 \\ &= (\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda \|\mathbf{w}\|_2^2 \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{w}^T(\mathbf{X}^T\mathbf{y}) + \mathbf{w}^T(\mathbf{X}^T\mathbf{X})\mathbf{w} + \lambda \mathbf{w}^T\mathbf{w} \\ &= \mathbf{y}^T\mathbf{y} - 2\mathbf{w}^T(\mathbf{X}^T\mathbf{y}) + \mathbf{w}^T(\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I}_{D\times D})\mathbf{w} \end{aligned}$$

Optimization (analytical/closed-form solution):
$\nabla J(\mathbf{w}) = 0 \Rightarrow -2\mathbf{X}^T\mathbf{y} + 2(\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I}_{D\times D})\mathbf{w} = 0 \Rightarrow \mathbf{w}^* = (\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I}_{D\times D})^{-1}\mathbf{X}^T\mathbf{y}$

Logistic Regression
Model: $f : \mathbf{x} \rightarrow y, \; f(\mathbf{x}) = \begin{cases} 1, & \sigma(\mathbf{w}^T\mathbf{x}) > 0.5 \\ 0, & \text{otherwise} \end{cases}, \;\; \sigma(\eta) = \frac{1}{1+e^{-\eta}}$

Training data: $\mathcal{D} = \{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_N}, y_N)\}$

Evaluation through cross-entropy error (no regularization):
$\mathcal{E}(\mathbf{w}) = -\sum_{n=1}^{N} \left\{ y_n \log\left[\sigma(\mathbf{w}^T\mathbf{x_n})\right] + (1 - y_n)\log\left[1 - \sigma(\mathbf{w}^T\mathbf{x_n})\right] \right\}$

Evaluation through cross-entropy error ($l2$-norm regularization):
$\mathcal{E}(\mathbf{w}) = -\sum_{n=1}^{N} \left\{ y_n \log\left[\sigma(\mathbf{w}^T\mathbf{x_n})\right] + (1 - y_n)\log\left[1 - \sigma(\mathbf{w}^T\mathbf{x_n})\right] \right\} + \lambda \mathbf{w}^T\mathbf{w}$

$$\begin{aligned} \nabla \mathcal{E}(\mathbf{w}) &= \sum_{n=1}^{N} \left(\sigma(\mathbf{w}^T\mathbf{x_n}) - y_n\right) \mathbf{x_n} + 2\lambda \mathbf{w} \\ \mathbf{H} &= \nabla \left((\nabla \mathcal{E}(\mathbf{w}))^T \right) \\ &= \nabla \left(\sum_{n=1}^{N} \left(\sigma(\mathbf{w}^T\mathbf{x_n}) - y_n\right) \mathbf{x_n}^T + 2\lambda \mathbf{w}^T \right) \\ &= \sum_{n=1}^{N} \underbrace{\sigma(\mathbf{w}^T\mathbf{x_n})}_{\in[0,1]} \cdot \underbrace{\left(1 - \sigma(\mathbf{w}^T\mathbf{x_n})\right)}_{\in[0,1]} \cdot \underbrace{\left(\mathbf{x_n} \cdot \mathbf{x_n}^T\right)}_{\in\mathcal{R}^{D\times D}} + \lambda \mathbf{I}_{D\times D} \end{aligned}$$